When is the rate function of a random vector strictly convex?

Article (Published Version)

Electron. Commun. Probab. **26** (2021), article no. 41, 1–11.
https://doi.org/10.1214/21-ECP409
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

When is the rate function of a random vector strictly convex?*

Vladislav Vysotsky†

Abstract

We give a necessary and sufficient condition for strict convexity of the rate function of a random vector in $\mathbb{R}^d$. This condition is always satisfied when the random vector has finite Laplace transform. We also completely describe the effective domain of the rate function under a weaker condition.

Keywords: rate function; convex conjugate; Legendre–Fenchel transform; strictly convex; strict convexity; effective domain; steep; essentially smooth; essentially strictly convex.
MSC2020 subject classifications: Primary 60E10; 26B25, Secondary 60F10.
Submitted to ECP on September 16, 2020, final version accepted on June 15, 2021.
Supersedes arXiv:2009.06809v2.

1 Introduction

Let X be a random vector in $\mathbb{R}^d$ and I_X be its *rate function*, given by

$$I_X(v) := \sup_{u \in \mathbb{R}^d} \big(u \cdot v - \log \mathbb{E} e^{u \cdot X} \big), \qquad v \in \mathbb{R}^d,$$

where '$\cdot$' stands for the scalar product in $\mathbb{R}^d$. This function is the *convex conjugate* of the *logarithmic Laplace transform* of X, defined by $K_X(u) := \log \mathbb{E} e^{u \cdot X}$ for every $u \in \mathbb{R}^d$.

The function K_X takes values in $(-\infty, +\infty]$, satisfies $K(0) = 0$, and is convex by Hölder's inequality. Then I_X is also convex and finite at least at one point ([8, Theorem 12.2]), and it takes values in $[0, +\infty]$. The *effective domain* $\mathcal{D}(I_X)$ of I_X, defined by

$$\mathcal{D}(I_X) := \{ v \in \mathbb{R}^d : I_X(v) < +\infty \},$$

is convex and non-empty, and so is the effective domain $\mathcal{D}(K_X)$ of K_X.

It is easy to show that K_X is differentiable at every point of $\operatorname{int} \mathcal{D}(K_X)$ ([2, Corollary 7.1]). When the set $\operatorname{int} \mathcal{D}(K_X)$ is non-empty, we say that K_X is *steep* (at the boundary of its effective domain) if $\lim_{n \to \infty} |\nabla K_X(u_n)| = \infty$ for every sequence $u_1, u_2, \dots$ in $\operatorname{int} \mathcal{D}(K_X)$ converging to a point in $\partial \mathcal{D}(K_X)$. Note that K_X is steep when it is finite at every point.

The property of steepness appears in a number of general convex-analytic results concerning the so-called *essentially smooth* convex functions on $\mathbb{R}^d$ ([8, Section 26]). In the context of large deviations probabilities this property features in the important

*This work was supported in part by Dr Perry James (Jim) Browne Research Centre.
†University of Sussex, United Kingdom. E-mail: v.vysotskiy@sussex.ac.uk

Gärtner–Ellis theorem ([4, Section 2.3]). The assumption $0 \in \operatorname{int} \mathcal{D}(K_X)$, which is of course stronger than $\operatorname{int} \mathcal{D}(K_X) \neq \varnothing$, is crucial for classical Cramér's theorem ([7, Section 2.4]), where the rate function I_X plays the key role.

Let us recall a few more definitions. For any $A \subset \mathbb{R}^d$, denote by $\operatorname{conv} A$ (resp. $\operatorname{aff} A$) the *convex hull* (resp. *affine hull*) of A, i.e. the minimal convex (resp. affine) subset of $\mathbb{R}^d$ containing A; denote by $\operatorname{rint} A$ is the *relative interior* of A, i.e. the interior of A in the relative topology of $\operatorname{aff} A$; and let $\partial_{\mathrm{rel}} A := \operatorname{cl} A \setminus \operatorname{rint} A$ be the *relative boundary* of A. Note that $\operatorname{rint} A = A$ if A consists of a single point.

The *topological support* of a random vector X in $\mathbb{R}^d$, denoted by $\operatorname{supp} X$, is the minimal by inclusion closed set $S \subset \mathbb{R}^d$ such that $\mathbb{P}(X \in S) = 1$. The *convex support* of X is

$$C_X := \operatorname{conv}(\operatorname{supp} X).$$

A hyperplane $L \subset \mathbb{R}^d$ *supports* a convex set $C \subset \mathbb{R}^d$ if L intersects C and C is contained in either of the two half-spaces of $\mathbb{R}^d$ that have L as their boundary ($C \subset L$ is possible).

We say that I_X is *strictly convex on a set* $A \subset \mathcal{D}(I_X)$ if I_X is affine on no line segment contained in A, and I_X is *strictly convex* if it is strictly convex on $\mathcal{D}(I_X)$.

Our starting point is the following assertion.

Proposition 1.1. *Let X be a random vector in $\mathbb{R}^d$, $d \geq 1$.*

a) We have

$$\operatorname{rint} C_X \subset \mathcal{D}(I_X) \subset \operatorname{cl} C_X, \tag{1.1}$$

hence $\operatorname{rint} \mathcal{D}(I_X) = \operatorname{rint} C_X$. Moreover, $I_X = +\infty$ on every hyperplane L in $\mathbb{R}^d$ supporting $\operatorname{cl} C_X$ and such that $\mathbb{P}(X \in L) = 0$.

b) I_X is strictly convex on $\operatorname{rint} \mathcal{D}(I_X)$ if and only if $\operatorname{int} \mathcal{D}(K_X) \neq \varnothing$ and K_X is steep.

We will prove this result in full for completeness of exposition. The inclusions in (1.1) are not new but we do not have exact references. They are stated in [2, Theorem 9.1], which however concerns only a specific type of distributions. They follow from [3, Theorems 2.1, 2.3, 3.2] but do not appear in [3] explicitly. The last claim of Part a) is in [2, Theorem 9.5]. Part b) states a particular case of a general convex-analytic result [8, Theorem 26.3], with the novelty that we strengthened the direct implication.

The main result of this note is a necessary and sufficient condition (see Theorem 2.11) for strict convexity of I_X on its whole effective domain $\mathcal{D}(I_X)$. This condition is always satisfied when the Laplace transform of X is finite in the whole of $\mathbb{R}^d$, and thus I_X is strictly convex for such X.

In view of Proposition 1.1, we only need to characterize strict convexity of the rate function on the relative boundary of $\mathcal{D}(I_X)$. Our approach is based on the following result.

Theorem 1.2. *Let X be a random vector in $\mathbb{R}^d$, $d \geq 1$, and L be a hyperplane in $\mathbb{R}^d$ supporting C_X and such that $0 < \mathbb{P}(X \in L) < 1$. Then*

$$I_X(v) = I_{X|L}(v) - \log \mathbb{P}(X \in L), \qquad v \in L, \tag{1.2}$$

if and only if

$$\operatorname{Pr}_L(\operatorname{rint} \mathcal{D}(K_X)) = \operatorname{Pr}_L(\operatorname{rint} \mathcal{D}(K_{X|L})), \tag{1.3}$$

where $X|L$ is a random vector distributed as X conditioned to be in L and Pr_L denotes the orthogonal projection from $\mathbb{R}^d$ onto L.

Assume that condition (1.3) is satisfied for every hyperplane L as above. Then we can apply Proposition 1.1 to each of the $I_{X|L}$'s. By (1.2), this ensures that $\operatorname{rint} \mathcal{D}(C_{X|L}) \subset \mathcal{D}(I_X)$, and I_X is strictly convex on every set $\operatorname{rint} \mathcal{D}(C_{X|L})$ if and only if the $K_{X|L}$'s are steep. The main idea is to apply this argument to the $X|L$'s and further on *recursively*, using that the sets $\operatorname{rint} \mathcal{D}(I_{X|L})$ are disjoint with $\operatorname{rint} \mathcal{D}(I_X)$ by $\dim C_{X|L} < \dim C_X$. Under appropriate conditions, which ensure that (1.3) is satisfied at every step of the recursion, this lets us fully describe $\mathcal{D}(I_X)$ (see Corollary 2.13) and characterize strict convexity of I_X (see Theorem 2.11). We give the details in the next section, where we also comment on condition (1.3) (see Remark 2.6).

The property of strict convexity can be useful when proving uniqueness of solutions to minimization problems involving I_X. Such problems arise from large deviations principles, most naturally in Cramér's theorem (see [7, Section 2.4]) on random walks in $\mathbb{R}^d$. There are functional versions of this result, which describe scaled trajectories of random walks and continuous time analogues for Lévy processes (see [4, Section 5.1 and 5.2] and [6]). In both cases, if the increments have finite Laplace transform, then the large deviations are described by the rate function I of the form $I(f) = \int_0^1 I_X(f'(t))dt$ for f in AC_0, the space of coordinate-wise absolutely continuous $\mathbb{R}^d$-valued functions on $[0, 1]$ such that $f(0) = 0$.

For concrete examples, let $(S_n)_{n\in\mathbb{N}}$ be a random walk with i.i.d. increments distributed as X. When I_X is strictly convex, its unique minimizer b over a convex Borel set $B \subset \mathbb{R}^d$ that meets $\mathcal{D}(I_X)$ can be interpreted as the limit constant in the law of large numbers for the averages S_n/n conditioned to be in B. Under this conditioning, a typical trajectory $(S_k/n)_{1\le k\le n}$ of the random walk with $\mathcal{D}(K_X) = \mathbb{R}^d$ is asymptotically linear with slope b because the function $f_0(t) = bt$ is the unique minimizer of I over the set $\{f \in AC_0 : f(1) \in B\}$. This follows from Jensen's inequality using that I_X is strictly convex (by Corollary 2.12). When $0 \in \operatorname{int} \mathcal{D}(K_X)$ but $\mathcal{D}(K_X) \neq \mathbb{R}^d$, the rate function I has a more complicated form, and without strict convexity of I_X the argument above becomes less simple (see [5, pp. 16-17]). More elaborate examples arise, e.g. in the study [1] of large deviations of the perimeter and the area of convex hulls of planar random walks, where strict convexity of I_X simplified considerations.

Finally, we note that relating the conditional limit laws to the minimizers of the rate function, as above, corresponds to the fundamental *Gibbs conditioning principle* of statistical mechanics (see [4], including Sections 3.3 and 7.3).

2 Main result

We first recall some facts on the structure of convex sets.

A *face* of a non-empty convex set $C \subset \mathbb{R}^d$ is a convex subset C' of C such that every closed line segment in C with a relative interior point in C' has both endpoints in C'. Note that C itself is a face; the zero-dimensional faces are called the *extreme points* of C. If $L \subset \mathbb{R}^d$ is a hyperplane supporting C, then $C \cap L$ is face of C. Every face of such form is called *exposed*.

Denote by $\mathcal{F}(C)$ the set of non-empty faces of C and by $\mathcal{F}^*(C)$ its subset of *maximal proper faces*, defined by

$$\mathcal{F}^*(C) := \big\{C' \in \mathcal{F}(C) \setminus \{C\} : C' \not\subset C'' \text{ for every } C'' \in \mathcal{F}(C) \setminus \{C, C'\}\big\}.$$

We will use extensively that every face in $\mathcal{F}^*(C)$ is exposed (this follows from [8, Theorem 11.6 and Corollary 18.1.3]). Our need in the set $\mathcal{F}^*(C)$ is due to the following result.

Lemma 2.1. *Let $C \subset \mathbb{R}^n$ be a non-empty convex set. Then*

$$C \setminus \operatorname{rint} C = \bigcup_{C' \in \mathcal{F}^*(C)} C'. \tag{2.1}$$

Proof. In fact, the set $C \setminus \operatorname{rint} C$ contains every proper face of C by [8, Corollary 18.1.3]. On the other hand, by [8, Theorem 11.6], for every point in $C \setminus \operatorname{rint} C$ there is a hyperplane L containing this point and supporting C but not containing C. Then $C \cap L$ is a proper face of C. To finish the proof it remains to argue that for any proper face C' of C is contained in a maximal proper face of C.

Let us use induction in $\dim C'$. In the base case $\dim C' = \dim C - 1$, we always have $C' \in \mathcal{F}^*(C)$. Indeed, if this were not true, there would be a proper face C'' of C other than C' that strictly contains C'. Then C' would be a face of C'' (by definition of a face), hence $\dim C' < \dim C''$ by [8, Corollary 18.1.3]. This is a contradiction because there are no faces of C other than itself of dimension $\dim C$.

Let us prove the inductive step. If C' is maximal, we are done. Otherwise, choose C'' as above. By the assumption of induction, there is a $C''' \in \mathcal{F}^*(C)$ containing C''. This is a face required. □

We now consider faces of the convex support C_X of a random vector X in $\mathbb{R}^d$. First note that C_X is not necessarily closed; it can be even open.

Example 2.2. Let X be a random vector in $\mathbb{R}^2$ such that $\operatorname{supp} X = \{(x, y) \in \mathbb{R}^2 : y \geq \frac{1}{1+x^2}\}$. Then C_X is the open upper half-plane.

However, we have the following measurability result.

Lemma 2.3. *Let X be a random vector in $\mathbb{R}^d$, $d \geq 1$. Then C_X is a Borel subset of $\mathbb{R}^d$, and so is every $C \in \mathcal{F}^*(C_X)$.*

Proof. By Carathéodory's theorem ([8, Theorem 17.1]), every point in C_X is a convex combination of $d+1$ points in $\operatorname{supp} X$. Then $C_X = \cup_{n=1}^{\infty} \operatorname{conv}\big(\operatorname{supp} X \cap \{u \in \mathbb{R}^d : \|u\| \leq n\}\big)$. By [8, Theorem 17.2], each set under the union is closed, and hence C_X is Borel.

Every $C \in \mathcal{F}^*(C_X)$ is an exposed face of C_X, therefore $C = C_X \cap L$ for some affine hyperplane L supporting C_X. Hence C also is a Borel set. □

The lemma ensures that the following set is well-defined:

$$\mathcal{F}_+^*(C_X) := \{C \in \mathcal{F}^*(C_X) : \mathbb{P}(X \in C) > 0\}.$$

In the results below it is useful to know when this set is empty. We give the following criterion.

Lemma 2.4. *Let X be a random vector in $\mathbb{R}^d$, $d \geq 1$. Then $\mathcal{F}_+^*(C_X)$ is empty if and only if there is no hyperplane L in $\mathbb{R}^d$ supporting C_X and such that $0 < \mathbb{P}(X \in L) < 1$.*

Proof. If L is a hyperplane supporting C_X, then either $C_X \subset L$, in which case $\mathbb{P}(X \in L) = 1$, or $C_X \cap L \in \mathcal{F}^*(C_X)$, hence from $\mathcal{F}_+^*(C_X) = \varnothing$ we get $\mathbb{P}(X \in C_X \cap L) = 0$ and thus $\mathbb{P}(X \in L) = 0$. This proves the direct implication.

To prove the reverse implication, assume that there is a $C \in \mathcal{F}_+^*(C_X)$. This is an exposed face of C_X, therefore $C = C_X \cap L$ for some affine hyperplane L supporting C_X. Then $\mathbb{P}(X \in L) = \mathbb{P}(X \in C) > 0$, hence by the assumption, it must be $\mathbb{P}(X \in L) = 1$. Hence $\operatorname{supp} X \subset L$ (because L is a closed set) and therefore $C_X \subset L$. Thus, C is not a proper face of C_X, which is a contradiction. □

For every random vector X in $\mathbb{R}^d$ and $C \in \mathcal{F}_+^*(C_X)$, let $X|C$ be a random vector distributed as X conditioned on $X \in C$.

We now give two key definitions, both having recursive structure.

Definition 2.5. *We say that K_X has the projection property*[1] *if*

a) for every hyperplane L in $\mathbb{R}^d$ supporting C_X and such that $0 < \mathbb{P}(X \in L) < 1$, we have

$$\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_X)) = \mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_{X|L}));$$

b) $K_{X|C}$ has the projection property for every $C \in \mathcal{F}_+^*(C_X)$.

The projection property is well-defined since the definition allows us to identify, using recursion in $\dim C_X$, whether each particular K_X has this property nor not. This is true because 1) $\dim C_{X|C} \le \dim C < \dim C_X$ for every $C \in \mathcal{F}_+^*(C_X)$; 2) the recursion terminates (confirming that K_X has the property) if Conditions a) and b) hold vacuously, namely when $\mathcal{F}_+^*(C_X) = \varnothing$ (by Lemma 2.4); and 3) the recursion always terminates since $\mathcal{F}_+^*(C_X) = \varnothing$ when $\dim C_X = 0$, i.e. X is constant a.s.

Remark 2.6. Let us comment on Condition a).

a) Each set $\mathcal{D}(K_{X|L})$ is a right cylinder. So is its relative interior, which satisfies

$$\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_{X|L})) = \mathrm{rint}(\mathrm{Pr}_L\,\mathcal{D}(K_{X|L})) = \mathrm{rint}(L \cap \mathcal{D}(K_{X|L})) = L \cap \mathrm{rint}\,\mathcal{D}(K_{X|L}),$$

where the first and the last equalities follow from [8, Theorem 6.6 and Corollary 6.5.1].

b) We always have $\mathcal{D}(K_X) \subset \mathcal{D}(K_{X|L})$. This follows from

$$K_X(u) \ge \log \mathbb{E}\big(e^{u\cdot X}\mathbb{1}_{\{X\in L\}}\big) = K_{X|L}(u) + \log \mathbb{P}(X \in L), \qquad u \in \mathbb{R}^d. \tag{2.2}$$

Assume additionally that $\mathcal{D}(K_X)$ is not entirely contained in the relative boundary of $\mathcal{D}(K_{X|L})$; this holds, e.g. when $\mathrm{int}\,\mathcal{D}(K_X) \ne \varnothing$ or $0 \in \mathrm{rint}\,\mathcal{D}(K_{X|L})$. Then

$$\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_X)) \subset \mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_{X|L})) \tag{2.3}$$

because $\mathrm{rint}\,\mathcal{D}(K_X) \subset \mathrm{rint}\,\mathcal{D}(K_{X|L})$ by [8, Corollary 6.5.2]. Thus, (1.3) means that the projection of $\mathrm{rint}\,\mathcal{D}(K_X)$ on L does not increase if X is replaced by $X|L$.

c) Every supporting hyperplane L to C_X is of the form $L = \{v \in \mathbb{R}^d : \ell \cdot v = h_{C_X}(\ell)\}$, where $\ell \in \mathbb{S}^{d-1}$ is a unit vector orthogonal to L and h_{C_X} is the *support function* of C_X defined by $h_{C_X}(u) := \sup_{v\in C_X} u \cdot v$, $u \in \mathbb{R}^d$. Since $\ell \in \mathcal{D}(h_{C_X})$ if and only if C_X is bounded in direction ℓ (equivalently, $\mathrm{supp}\,X$ is bounded in direction ℓ), this implies that $\{a\ell : a \ge 0\} \subset \mathcal{D}(K_X)$. It is therefore easy to see that $\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_X)) = L$ when $\ell \in \mathrm{int}\,\mathcal{D}(h_{C_X})$. Hence for such ℓ equality (1.3) always holds true by $\mathcal{D}(K_X) \subset \mathcal{D}(K_{X|L})$.

Thus, it suffices to check the assumption of Condition a) only for hyperplanes supporting C_X that are orthogonal to directions in the set $\partial\mathcal{D}(h_{C_X}) \cap \mathbb{S}^{d-1}$. For $d = 2$ this set contains at most two directions because $\mathcal{D}(h_{C_X})$ is a convex cone.

We now give a few examples.

Example 2.7. K_X has the projection property in the following cases:

a) $\mathcal{F}_+^*(C_X)$ is empty. In particular, this holds true when $\mathbb{P}(X \in \partial_{\mathrm{rel}} C_X) = 0$; see (2.1).

b) $\mathcal{D}(K_X) = \mathbb{R}^d$ or, equivalently, $\mathbb{E}e^{u\cdot X} < \infty$ for every $u \in \mathbb{R}^d$; cf. (2.2).

c) $d = 1$.

[1] Strictly speaking, this is a property of the distribution of X rather than of K_X. However, the distributions that satisfy $\mathrm{int}\,\mathcal{D}(K_X) \ne \varnothing$ are determined by their Laplace transform (this reduces to $d = 1$, where Theorem 6a in Chapter VI of [9] applies). Note that our main result, Theorem 2.11.b, assumes $\mathrm{int}\,\mathcal{D}(K_X) \ne \varnothing$.

d) $d = 2$ and equality (1.3) holds true for every line L of the form $L = \operatorname{aff} C$, where $C \in \mathcal{F}_+^*(C_X)$ is unbounded (there are at most two such faces).

Indeed, such lines are orthogonal to the directions in $\partial\mathcal{D}(h_{C_X}) \cap \mathbb{S}^1$ and then Remark 2.6.c applies. Clearly, Condition a) in Definition 2.5 is satisfied by Example 2.7.c since $\dim C_{X|C} \le 1$ for every $C \in \mathcal{F}_+^*(C_X)$.

Our second key definition is as follows.

Definition 2.8. *If $\mathcal{D}(K_X)$ has non-empty interior, we say that K_X is totally steep if K_X is steep and $K_{X|C}$ is totally steep for every $C \in \mathcal{F}_+^*(C_X)$.*

Again, this property is well-defined by recursion in $\dim C_X$ because 1) $\dim C_{X|C} < \dim C_X$ for every $C \in \mathcal{F}_+^*(C_X)$; 2) $\operatorname{int}\mathcal{D}(K_{X|C}) \ne \varnothing$ for $C \in \mathcal{F}_+^*(C_X)$ by $\mathcal{D}(K_X) \subset \mathcal{D}(K_{X|C})$ (cf. (2.2)); 3) K_X is totally steep when it is steep and $\mathcal{F}_+^*(C_X) = \varnothing$; and 4) K_X is totally steep when $\dim C_X = 0$.

Example 2.9. K_X is totally steep if $\mathcal{D}(K_X) = \mathbb{R}^d$.

Example 2.10. Let us construct K_X which neither has the projection property nor is totally steep. Put $X := (\alpha X_1, \alpha X_2 + (1-\alpha)X_3)$, where X_1, X_2, X_3, α are independent non-negative random variables such that X_1 and X_2 have the standard exponential distribution with density e^{-x} for $x > 0$, X_3 has the absolutely continuous distribution with density proportional to $e^{-2x}/(1+x^3)$ for $x > 0$, and $\mathbb{P}(\alpha = 0) = \mathbb{P}(\alpha = 1) = 1/2$.

We have $\mathcal{D}(K_{X_3}) = (-\infty, 2]$ and it is easy to check that $K'_{X_3}(2-) < +\infty$, hence K_{X_3} is not steep; and K_{X_1} is steep. Furthermore, $K_X(u_1, u_2) = \frac{1}{2}K_{X_1}(u_1) + \frac{1}{2}K_{X_2}(u_2) + \frac{1}{2}K_{X_3}(u_2)$ for $u_1, u_2 \in \mathbb{R}$; the set C_X is the closed positive quadrant in the plane; and $\mathcal{F}_+^*(C_X) = \{C\}$ with $C := \{0\} \times [0, \infty)$. We can see that K_X is steep but not totally steep because $\mathcal{D}(K_X) = (-\infty, 1) \times (-\infty, 1)$ but for the ordinate line $L = \operatorname{aff} C$ supporting C_X, the random vector $X|L$ is distributed as $(0, X_3)$ and thus $K_{X|L}$ is not steep. This also shows that Condition a) in Definition 2.5 is violated because $\operatorname{Pr}_L(\operatorname{rint}\mathcal{D}(K_{X|L})) = \{0\} \times (-\infty, 2)$ but $\operatorname{Pr}_L(\operatorname{rint}\mathcal{D}(K_X)) = \{0\} \times (-\infty, 1)$, and thus K does not have the projection property.

We are now ready to state the main result of the paper.

Theorem 2.11. *Let X be a random vector in $\mathbb{R}^d$, $d \ge 1$.*

a) If K_X satisfies Condition a) in Definition 2.5 of the projection property, then

$$\mathcal{F}^*(\mathcal{D}(I_X)) \subset \{\mathcal{D}(I_{X|C}) : C \in \mathcal{F}_+^*(C_X)\} \subset \mathcal{F}(\mathcal{D}(I_X)) \setminus \{\mathcal{D}(I_X)\}. \tag{2.4}$$

b) I_X is strictly convex if and only if $\operatorname{int}\mathcal{D}(K_X) \ne \varnothing$, K_X has the projection property, and K_X is totally steep.

Let us present a few corollaries.

Corollary 2.12. *If $\mathbb{E}e^{u \cdot X} < \infty$ for every $u \in \mathbb{R}^d$, then I_X is strictly convex.*

Proof. This follows directly from Part b) using Examples 2.7.b and 2.9. □

Corollary 2.13. *If K_X has the projection property, then $\mathcal{D}(I_X) \subset C_X$ and*

$$\begin{aligned}&\mathcal{D}(I_X)\\ &= \operatorname{rint} C_X \cup \bigcup_{C_1 \in \mathcal{F}_+^*(C_X)} \operatorname{rint} C_{X|C_1} \cup \bigcup_{C_2 \in \mathcal{F}_+^*(C_{X|C_1})} \operatorname{rint} C_{X|C_2} \cup \ldots \cup \bigcup_{C_d \in \mathcal{F}_+^*(C_{X|C_{d-1}})} \operatorname{rint} C_{X|C_d}.\end{aligned}$$

Proof. We have

$$\mathcal{D}(I_X) = \operatorname{rint}\mathcal{D}(I_X) \cup \bigcup_{C \in \mathcal{F}^*(\mathcal{D}(I_X))} C = \operatorname{rint} C_X \cup \bigcup_{C_1 \in \mathcal{F}_+^*(C_X)} \mathcal{D}(I_{X|C_1}),$$

where the first equality follows from (2.1) and the second one follows from (2.4) and the fact that $\operatorname{rint}\mathcal{D}(I_X) = \operatorname{rint} C_X$ (see Proposition 1.1.a). Then we establish the equality claimed by simple induction in $\dim C_X$ using that each random vector $(X|C_1)|C_2$ has the same distribution as $X|C_2$. In the base case $\dim C_X = 0$ the claim holds by $\mathcal{D}(I_X) = C_X = \operatorname{rint} C_X$ and $\mathcal{F}^*_+(C_X) = \varnothing$. The same inductive argument establishes the inclusion $\mathcal{D}(I_X) \subset C_X$. □

Corollary 2.14. *Assume that K_X has the projection property. Then v is an extreme point of $\mathcal{D}(I_X)$ if and only if v is an extreme point of C_X and $\mathbb{P}(X = v) > 0$. For such v, we have $I_X(v) = -\log \mathbb{P}(X = v)$.*

Proof. Assume that v is an extreme point of C_X and $\mathbb{P}(X = v) > 0$. We use induction in $\dim C_X$. In the base case $\dim C_X = 0$, we simply have $I_X(v) = 0 = -\log(X = v)$. To prove the induction step for $\dim C_X \geq 1$, use that by (2.1) there is a face $C \in \mathcal{F}^*_+(C_X)$ that contains v. Then v is an extreme point of $C_{X|C}$ because $v \in C_{X|C}$ by $\mathbb{P}((X|C) = v) > 0$ and v is an extreme point of the convex set C_X which contains $C_{X|C}$.

Since C an exposed face of C_X, there is a hyperplane L supporting C_X such that $C = C_X \cap L$. Then $X|C$ and $X|L$ have the same distribution since $\mathbb{P}(X \in L \setminus C) = 0$, and by (1.2) and the assumption of induction we get

$$I_X(v) = I_{X|C}(v) - \log \mathbb{P}(X \in C) = -\log \mathbb{P}((X|C) = v) - \log \mathbb{P}(X \in C) = -\log \mathbb{P}(X = v).$$

Then $I_X(v) < \infty$, and thus $v \in \mathcal{D}(I_X)$. Hence v is an extreme point of $\mathcal{D}(I_X)$ because v is an extreme point of the convex set C_X which contains $\mathcal{D}(I_X)$ by Corollary 2.13.

Proving the reverse implication is similar. For the induction step, for $\dim \mathcal{D}(I_X) \geq 1$, use that by (2.1) there is a face $F \in \mathcal{F}^*(\mathcal{D}(I_X))$ that contains v. Then v is an extreme point of F. By (2.4), $F = \mathcal{D}(I_{X|C})$ for some face $C \in \mathcal{F}^*_+(C_X)$, and we can apply the assumption of induction as above. □

3 Proofs

Proof of Proposition 1.1. a) Recall that $C_X = \operatorname{conv}(\operatorname{supp} X)$. Fix a $v \notin \operatorname{cl} C_X$. By [8, Corollary 11.5.1], there exists a non-zero $u_0 \in \mathbb{R}^d$ such that $u_0 \cdot x < u_0 \cdot v$ for any $x \in \operatorname{supp} X$. In other words, $u_0 \cdot X < u_0 \cdot v$ a.s. By the monotone convergence theorem, we get

$$I(v) = \sup_{u \in \mathbb{R}^d} (u \cdot v - K_X(u)) \geq \sup_{a>0} (a u_0 \cdot v - K_X(a u_0)) = -\inf_{a>0} (\log \mathbb{E} e^{a(u_0 \cdot X - u_0 \cdot v)}) = +\infty. \tag{3.1}$$

Thus, $\mathcal{D}(I_X) \subset \operatorname{cl} C_X$.

Furthermore, if L a hyperplane supporting $\operatorname{cl} C_X$, take any non-zero $u_0 \in \mathbb{R}^d$ orthogonal to L and directed such that $u_0 \cdot x \leq u_0 \cdot v$ for any $x \in \operatorname{supp} X$ and $v \in L$, that is $u_0 \cdot X \leq u_0 \cdot v$ a.s. This inequality is strict if $\mathbb{P}(X \in L) = 0$, in which case $I(v) = +\infty$ holds true by (3.1), as required.

We now show that $\operatorname{rint} C_X \subset \mathcal{D}(I_X)$. Assume that this does not hold. Then, since $\mathcal{D}(I_X) \subset \operatorname{cl} C_X$ and the sets C_X and $\mathcal{D}(I_X)$ are convex, we have $\operatorname{cl}\mathcal{D}(I_X) \neq \operatorname{cl} C_X$ by [8, Corollary 6.3.1]. Therefore, there exists an open ball $B \subset \mathbb{R}^d$ such that $\operatorname{cl}\mathcal{D}(I_X) \cap \operatorname{cl} B = \varnothing$ and $\operatorname{cl} C_X \cap B \neq \varnothing$.

For any $n \in \mathbb{N}$, let S_n be the sum of n independent identically distributed copies of X. Then for any $u \in \mathcal{D}(K_X)$, we have

$$\mathbb{P}(S_n/n \in B) = \mathbb{E}[\mathbb{1}(S_n/n \in B)] \leq \mathbb{E}\big[\mathbb{1}\big(u \cdot S_n \geq n \inf_{v \in \operatorname{cl} B} u \cdot v\big)\big] \leq e^{-n \inf_{v \in \operatorname{cl} B} u \cdot v}\, \mathbb{E} e^{u \cdot S_n},$$

where the last equality follows from Markov's inequality. Then

$$n^{-1}\log\mathbb{P}(S_n/n\in B)\le \inf_{u\in\mathcal{D}(K_X)}\Big(-\inf_{v\in\operatorname{cl}B}(u\cdot v-K_X(u))\Big)=-\sup_{u\in\mathcal{D}(K_X)}\inf_{v\in\operatorname{cl}B}(u\cdot v-K_X(u)).$$

Finally, let us interchange the supremum and the infimum using a minimax result [8, Corollary 37.3.2] on concave-convex functions. This gives

$$n^{-1}\log\mathbb{P}(S_n/n\in B)\le -\inf_{v\in\operatorname{cl}B}I_X(v). \tag{3.2}$$

This inequality appears, e.g., in [7, Eq. (2.16)].

On the other hand, since $\operatorname{cl}C_X\cap B\ne\varnothing$, B is open, and C_X is convex, it follows from [8, Corollary 6.3.2] that $\operatorname{rint}C_X$ intersects with B. Hence, by Carathéodory's theorem ([8, Theorem 17.1]), there is a convex combination $\sum_{i=1}^m\alpha_i x_i\in B$, where m is a positive integer, $x_i\in\operatorname{supp}X$ and $\alpha_i>0$ for every $1\le i\le m$, and $\sum_{i=1}^m\alpha_i=1$. By finding a rational approximation to all but one of the α_i's, we get $\frac{1}{n}\sum_{i=1}^m n_i x_i\in B$ for some positive integer n_i and $n=\sum_{i=1}^m n_i$. Furthermore, there exist open balls $B_i\subset\mathbb{R}^d$ such that $x_i\in B_i$ for every $1\le i\le m$ and $\frac{1}{n}\sum_{i=1}^m n_iB_i\subset B$. Since each open ball B_i intersects with $\operatorname{supp}X$, we have $\mathbb{P}(X\in B_i)>0$. Therefore,

$$\mathbb{P}(S_n/n\in B)\ge \Pi_{i=1}^m\mathbb{P}(X\in B_i)^{n_i}>0. \tag{3.3}$$

Inequalities (3.2) and (3.3) imply that $\operatorname{cl}B\cap\mathcal{D}(I_X)\ne\varnothing$, which is contradiction. Thus, we proved that $\operatorname{rint}C_X\subset\mathcal{D}(I_X)\subset\operatorname{cl}C_X$, establishing (1.1). Finally, by [8, Corollary 6.3.1] this gives $\operatorname{rint}C_X=\operatorname{rint}\mathcal{D}(I_X)$, as required.

b) Put $d':=\dim C_X$. We assume that $d'\ge 1$, otherwise the claim is trivial.

Recall that I_X is *subdifferentiable* at a point $v_0\in\mathbb{R}^d$ if there is a $u\in\mathbb{R}^d$ such that the inequality $I_X(v)\ge I_X(v_0)+u\cdot(v-v_0)$ holds for every $v\in\mathbb{R}^d$. We claim that I_X is subdifferentiable at no point outside of $\operatorname{rint}\mathcal{D}(I_X)$. Combined with the fact that K_X is differentiable at every point of $\operatorname{int}\mathcal{D}(K_X)$ ([2, Corollary 7.1]), this implies that the asserted necessary and sufficient condition for strict convexity of I_X is a particular case of [8, Theorem 26.3].

Assume first that $d'=d$. In this case K_X is strictly convex by [2, Theorem 7.1]; this actually follows immediately from the criterion for equality in Hölder's inequality. Therefore, K_X is *essentially strictly convex*, i.e. K_X is strictly convex on every interval contained in the set of points where K_X is subdifferentiable. Hence I_X is *essentially smooth* by [8, Theorem 26.3], that is I_X is differentiable on the set $\operatorname{int}\mathcal{D}(I_X)$, which is required to be non-empty, and I_X is steep. By [8, Theorem 26.1], this implies that I_X is not subdifferentiable outside of $\operatorname{rint}\mathcal{D}(I_X)$, as required.

In the remaining case $1\le d'\le d-1$, put $L:=\operatorname{aff}(\operatorname{supp}X)$. We can assume w.l.o.g. that $0\in L$, otherwise pick any $\mu\in L$ and use the simple fact that $I_X(v)=I_{X-\mu}(v-\mu)$ for $v\in\mathbb{R}^d$, which easily implies that our claim holds true for I_X if and only if it holds for $I_{X-\mu}$.

Since L is a linear subspace of $\mathbb{R}^d$ of dimension d', there exists an orthogonal mapping $U:L\to\mathbb{R}^{d'}$. Then by $X\in L$ a.s., for any $v\in L$ we have

$$I_X(v)=\sup_{u\in\mathbb{R}^d}\big(u\cdot v-\log\mathbb{E}e^{u\cdot X}\big)=\sup_{u\in L}\big(u\cdot v-\log\mathbb{E}e^{u\cdot X}\big)=I_{U(X)}(U(v)), \tag{3.4}$$

where in the last equality we used the change of variables $u\mapsto U(u)$. Therefore, since the mapping U is linear and invertible, I_X is subdifferentiable at a $v\in L$ if and only if $I_{U(X)}$ is subdifferentiable at $U(v)$ by [8, Theorem 23.9] (applied with $f=I_{U(X)}$ and $A=U^{-1}$). On the other hand, $v\in\operatorname{rint}\mathcal{D}(I_X)$ if and only if $U(v)\in\operatorname{rint}\mathcal{D}_{U(X)}$, since $U(\operatorname{rint}\mathcal{D}(I_X))=\operatorname{rint}U(\mathcal{D}(I_X))=\operatorname{rint}\mathcal{D}(I_{U(X)})$ by [8, Theorem 6.6]. Thus, by equality (3.4), the case $d'<d$ reduces to the case $d'=d$ because the support of the random vector $U(X)$ in $\mathbb{R}^{d'}$ has full dimension. This finishes the proof of the claim. □

Our proofs of Theorems 1.2 and 2.11 rely on the following technical result, where * stands for convex conjugation (the Legendre–Fenchel transform) of functions on $\mathbb{R}^d$.

Lemma 3.1. *Let X be a random vector in $\mathbb{R}^d$, $d \ge 1$, and L be a hyperplane in $\mathbb{R}^d$ supporting C_X and such that $\mathbb{P}(X \in L) > 0$. Put $\tilde{K}_{X|L}(u) := K_{X|L}(u)$ if $u \in \mathrm{Pr}_L^{-1}(\mathrm{Pr}_L \mathcal{D}(K_X))$, otherwise $\tilde{K}_{X|L}(u) := +\infty$ for $u \in \mathbb{R}^d$. Then*

$$I_X(v) = (\tilde{K}_{X|L})^*(v) - \log \mathbb{P}(X \in L), \qquad v \in L, \tag{3.5}$$

and $(\tilde{K}_{X|L})^*(v) = +\infty$ for $v \notin L$. Moreover, we have

$$\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(\tilde{K}_{X|L})) = \mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_X)). \tag{3.6}$$

Proof. Denote by L_0 the hyperplane passing through 0 that is parallel to L, and let $\ell \in \mathbb{R}^d$ be the unit vector orthogonal to L_0 such that $\ell \cdot u \le \ell \cdot v$ for any $u \in C_X$ and $v \in L$. Denote by (v_1, v_2) the coordinates of $v \in L$ in $L_0 \oplus L^\perp$, where $L^\perp := \mathbb{R}\ell$.

For any $u_1 \in L_0$ such that $\mathbb{E}e^{(u_1+u_2\ell)\cdot X} < \infty$ for some real $u_2 = u_2'$, we have

$$\begin{aligned}\sup_{u_2 \in \mathbb{R}} \big(u_2 v_2 - \log \mathbb{E}e^{(u_1+u_2\ell)\cdot X}\big) &= -\log\big(\inf_{u_2\in\mathbb{R}} \mathbb{E}e^{u_1\cdot X + u_2(\ell\cdot X - v_2)}\big) \\ &= -\log \mathbb{E}[e^{u_1\cdot X}\mathbb{1}_{\{X\in L\}}],\end{aligned} \tag{3.7}$$

where the last equality follows from the dominated convergence theorem applied as $u_2 \to +\infty$ using that $e^{u_1\cdot X + u_2'(\ell\cdot X - v_2)}$ is an integrable majorant, which is true because the function $u_2 \mapsto e^{u_1\cdot X+u_2(\ell\cdot X-v_2)}$ is non-increasing a.s. by $\ell\cdot X \le v_2$ a.s. On the other hand, if $u_1 \in L_0$ is such that $\mathbb{E}e^{(u_1+u_2\ell)\cdot X} = \infty$ for every real u_2, then the l.h.s. of the first equality in (3.7) is $-\infty$. Therefore,

$$\begin{aligned} I_X(v) &= \sup_{u_1\in L_0}\sup_{u_2\in\mathbb{R}} \big(u_1\cdot v_1 + u_2 v_2 - \log \mathbb{E}e^{(u_1+u_2\ell)\cdot X}\big) \\ &= \sup_{\substack{u_1\in L_0:\\ (u_1+\mathbb{R}\ell)\cap\mathcal{D}(K_X)\ne\varnothing}} \big(u_1\cdot v_1 - \log \mathbb{E}[e^{u_1\cdot X}\mathbb{1}_{\{X\in L\}}]\big) \\ &= \sup_{u_1\in \mathrm{Pr}_{L_0}(\mathcal{D}(K_X))}\sup_{u_2\in\mathbb{R}} \big(u_1\cdot v_1 + u_2v_2 - \log\mathbb{E}\big[e^{(u_1+u_2\ell)\cdot(X|L)}\big]\big) - \log\mathbb{P}(X\in L) \\ &= \sup_{u\in \mathrm{Pr}_L^{-1}(\mathrm{Pr}_L\mathcal{D}(K_X))} \big(u\cdot v - \log \mathbb{E}e^{u\cdot(X|L)}\big) - \log\mathbb{P}(X\in L), \end{aligned}$$

where in the last equality we used that L_0 and L are parallel. This proves (3.5).

The claim $(\tilde{K}_{X|L})^*(v) = +\infty$ for $v \notin L$ follows exactly as in (3.1) using that $\tilde{K}_{X|L}(u_0) = K_{X|L}(u_0) < +\infty$ for any $u_0 \in L^\perp$ by $0 \in \mathcal{D}(K_X)$.

Lastly, it follows from (2.2) that $K_{X|L}(u) < +\infty$ when $u \in \mathrm{Pr}_L^{-1}(\mathrm{Pr}_L\mathcal{D}(K_X))$. Therefore, by the definition of $\tilde{K}_{X|L}$, we have $\mathcal{D}(\tilde{K}_{X|L}) = \mathrm{Pr}_L^{-1}(\mathrm{Pr}_L\mathcal{D}(K_X))$, and we obtain (3.6) interchanging Pr_L and rint by [8, Theorem 6.6] as follows:

$$\mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(\tilde{K}_{X|L})) = \mathrm{rint}(\mathrm{Pr}_L\,\mathcal{D}(\tilde{K}_{X|L})) = \mathrm{rint}(\mathrm{Pr}_L\,\mathcal{D}(K_X)) = \mathrm{Pr}_L(\mathrm{rint}\,\mathcal{D}(K_X)). \qquad \square$$

Proof of Theorem 1.2. Let L be a hyperplane supporting C_X and such that $\mathbb{P}(X\in L) > 0$. By Lemma 3.1 and Proposition 1.1.a applied to $X|L$, we have $I_{X|L}(v) = (K_{X|L})^*(v) = +\infty$ and $(\tilde{K}_{X|L})^*(v) = +\infty$ for $v \notin L$. Therefore, the functions $(\tilde{K}_{X|L})^*$ and $(K_{X|L})^*$ coincide if they are equal on L. Thus, (3.5) implies that

$$I_X(v) = I_{X|L}(v) - \log\mathbb{P}(X\in L), \qquad v \in L,$$

if and only if $(\tilde{K}_{X|L})^* = (K_{X|L})^*$. This is in turn equivalent to $(\tilde{K}_{X|L})^{**} = K_{X|L}$ (by [8, Theorem 12.2]) because $\tilde{K}_{X|L}$ is a convex function (this follows from the definition of $\tilde{K}_{X|L}$) and $K_{X|L}$ is a lower semi-continuous convex function (by [2, Theorem 7.1]), both finite at least at one point. The last equality holds true if and only if $\tilde{K}_{X|L}$ equals $K_{X|L}$ except possibly at some relative boundary points of $\mathcal{D}(\tilde{K}_{X|L})$ (by [8, Theorem 7.4]). Thus, equalities (1.2) and $\operatorname{rint}\mathcal{D}(\tilde{K}_{X|L}) = \operatorname{rint}\mathcal{D}(K_{X|L})$ are equivalent.

The latter one is equivalent to $\operatorname{Pr}_L(\operatorname{rint}\mathcal{D}(\tilde{K}_{X|L})) = \operatorname{Pr}_L(\operatorname{rint}\mathcal{D}(K_{X|L}))$ because the sets $\operatorname{rint}\mathcal{D}(\tilde{K}_{X|L})$ and $\operatorname{rint}\mathcal{D}(K_{X|L})$ are right cylinders by [8, Corollary 6.6.2]. Hence, by (3.6), equalities (1.2) and (1.3) are equivalent, as claimed. □

Proof of Theorem 2.11. a) Let us prove the first inclusion in (2.4). Let $F \in \mathcal{F}^*(\mathcal{D}(I_X))$ be a maximal proper face of $\mathcal{D}(I_X)$. Then there is a hyperplane L supporting the convex set $\mathcal{D}(I_X)$ such that $F = \mathcal{D}(I_X) \cap L$. The hyperplane L also supports $\operatorname{cl} C_X$ by the second inclusion in (1.1). Moreover, we have $\mathbb{P}(X \in L) > 0$ since otherwise $\mathcal{D}(I_X) \cap L = \varnothing$ by Proposition 1.1.a, which is a contradiction. Therefore, $L \cap C_X \neq \varnothing$, and thus L supports C_X. Hence $C := C_X \cap L$ is a face of C_X. We also have $F = \mathcal{D}(I_{X|C})$ by (1.2) and the fact that $X|C$ has the same distribution as $X|L$ (as $\mathbb{P}(X \in L \setminus C) = 0$). Hence $F = \mathcal{D}(I_X) \cap \operatorname{cl} C_{X|C}$ by (1.1).

Clearly, C is a proper face of C_X (i.e. $C \neq C_X$) since otherwise F cannot be a proper face of $\mathcal{D}(I_X)$. However, C is not necessarily a maximal proper face. Let $C' \in \mathcal{F}^*(C_X)$ be such that $C \subset C'$. Since this is an exposed face of C_X, there is a hyperplane L' supporting C_X and satisfying $C' = C_X \cap L'$. We have

$$\mathbb{P}(X \in L') = \mathbb{P}(X \in C') \geq \mathbb{P}(X \in C) > 0.$$

Since L' supports C_X, equality (1.2) is valid with $L = L'$ and it implies that the set $F' := \mathcal{D}(I_X) \cap L'$ satisfies $F' = \mathcal{D}(I_{X|C'})$ and therefore is non-empty; moreover, we have $F' = \mathcal{D}(I_X) \cap \operatorname{cl} C_{X|C'}$ by (1.1). This shows that F' is a proper face of $\mathcal{D}(I_X)$ since L' supports $\mathcal{D}(I_X)$ by (1.1).

Finally, by $C \subset C'$, we have $C_{X|C} \subset C_{X|C'}$, and thus

$$F = \mathcal{D}(I_X) \cap \operatorname{cl} C_{X|C} \subset \mathcal{D}(I_X) \cap \operatorname{cl} C_{X|C'} = F'.$$

Therefore, $F = F'$ since F is a maximal proper face by the assumption. Thus, we have $F = \mathcal{D}(I_{X|C'})$, which proves the first inclusion in (2.4).

To prove the remaining inclusion in (2.4), pick a $C' \in \mathcal{F}^*_+(C_X)$. Then $C' = C_X \cap L'$ for some hyperplane L' supporting C_X and satisfying $\mathbb{P}(X \in L') > 0$. As we have shown just above, $F' := \mathcal{D}(I_X) \cap L'$ is a non-empty proper face of $\mathcal{D}(I_X)$ (but it is not necessarily a maximal one anymore) and $F' = \mathcal{D}(I_{X|C'})$. This finishes the proof of Part a).

b) Direct implication. Assume that I_X is strictly convex. Let us use induction in $\dim C_X$ to prove that $\operatorname{int}\mathcal{D}(K_X) \neq \varnothing$, K_X has the projection property, and K_X is totally steep.

This claim holds trivially in the base case $\dim C_X = 0$, where X is constant a.s.

To prove the induction step, consider any hyperplane L supporting C_X such that $0 < \mathbb{P}(X \in L) < 1$. By Lemma 3.1, the effective domain of the function $(\tilde{K}_{X|L})^*$ is contained in L. Therefore, this function is strictly convex by (3.5) because I_X is strictly convex by the assumption. Then $\operatorname{int}\mathcal{D}((\tilde{K}_{X|L})^{**}) \neq \varnothing$ and $(\tilde{K}_{X|L})^{**}$ is steep by [8, Theorem 26.3] (because strict convexity implies essential strict convexity). Hence $\operatorname{int}\mathcal{D}(\tilde{K}_{X|L}) \neq \varnothing$ and $\tilde{K}_{X|L}$ is also steep because it equals $(\tilde{K}_{X|L})^{**}$ except possibly at some relative boundary points of $\mathcal{D}(\tilde{K}_{X|L})$ ([8, Theorems 7.4 and 12.2]).

Let us show that $\operatorname{int}\mathcal{D}(\tilde{K}_{X|L}) = \operatorname{int}\mathcal{D}(K_{X|L})$. Otherwise, by $\mathcal{D}(\tilde{K}_{X|L}) \subset \mathcal{D}(K_{X|L})$ and convexity of $\mathcal{D}(K_{X|L})$, there is a point $u \in \partial\mathcal{D}(\tilde{K}_{X|L}) \cap \operatorname{int}\mathcal{D}(K_{X|L})$. Pick a sequence $u_1, u_2, \dots$ in $\operatorname{int}\mathcal{D}(\tilde{K}_{X|L})$ converging to u. Then $\lim_{n\to\infty} |\nabla \tilde{K}_{X|L}(u_n)| = \infty$ by steepness of $\tilde{K}_{X|L}$. Thus, $\lim_{n\to\infty} |\nabla K_{X|L}(u_n)| = \infty$ because $\tilde{K}_{X|L}$ equals $K_{X|L}$ whenever $\tilde{K}_{X|L} < \infty$, and hence $\tilde{K}_{X|L} = K_{X|L}$ on $\operatorname{int}\mathcal{D}(\tilde{K}_{X|L})$. However, it must be $\lim_{n\to\infty} |\nabla K_{X|L}(u_n)| =$

$|\nabla K_{X|L}(u)|$ because $K_{X|L}$ is continuously differentiable on $\operatorname{int}\mathcal{D}(K_{X|L})$ since so is the Laplace transform of any random variable ([2, Corollary 7.1]). This is a contradiction.

We now have $\mathrm{Pr}_L(\operatorname{int}\mathcal{D}(\tilde{K}_{X|L})) = \mathrm{Pr}_L(\operatorname{int}\mathcal{D}(K_{X|L}))$, which implies equality (1.3) by (3.6). Thus, K_X satisfies Condition a) in Definition 2.5 of the projection property because L was chosen arbitrarily. Equality (1.3) in turn implies (1.2) by Theorem 1.2, hence $I_{X|L}$ is strictly convex because so is I_X and $\mathcal{D}(I_{X|L}) \subset L$ by Proposition 1.1.a.

For any maximal proper face $C \in \mathcal{F}_+^*(C_X)$, pick a hyperplane L supporting C_X such that $C = C_X \cap L$. Since $0 < \mathbb{P}(X \in L) < 1$ and $\dim(\operatorname{supp}(X|L)) < \dim(\operatorname{supp} X)$, we can apply the assumption of induction to the random vector $X|L$, which is distributed as $X|C$ and has strictly convex rate function $I_{X|L}$ by the above. Therefore, $\operatorname{int}\mathcal{D}(K_{X|C}) \neq \varnothing$, $K_{X|C}$ has the projection property, and $K_{X|C}$ is totally steep. Thus, since C was chosen arbitrarily, K_X has the projection property, as required. Finally, K_X is totally steep, as required, since $\operatorname{int}\mathcal{D}(K_X) \neq \varnothing$ and K_X is steep by Proposition 1.1.b.

Reverse implication. Assume that $\varnothing \neq \operatorname{int}\mathcal{D}(K_X)$, K_X has the projection property, and K_X is totally steep. We again use induction in $\dim C_X$ to show that I_X is strictly convex.

In the base case $\dim C_X = 0$, the set $\mathcal{D}(I_X)$ consists of a single point, and the claim holds vacuously.

To prove the induction step, pick a closed line segment $J \subset \mathcal{D}(I_X)$. From (2.1) and the definition of a face, either $\operatorname{rint} J \subset \operatorname{rint}\mathcal{D}(I_X)$ or J is contained in some maximal proper face of $\mathcal{D}(I_X)$. In the former case, I_X is not affine on J by Proposition 1.1.b. In the latter case, by the first inclusion in (2.4), we have $J \subset \mathcal{D}(I_{X|C}) \subset C$ for some face $C \in \mathcal{F}_+^*(C_X)$. Note that $\varnothing \neq \operatorname{int}\mathcal{D}(K_X) \subset \operatorname{int}\mathcal{D}(K_{X|C})$ (cf. (2.2)), $K_{X|C}$ is totally steep since so is K_X, and $K_{X|C}$ has the projection property since K_X has this property. Therefore, $I_{X|C}$ is strictly convex by $\dim(\operatorname{supp}(X|C)) < \dim(\operatorname{supp} X)$ and the assumption of induction. Hence I_X is not affine on J because for some hyperplane L supporting C_X and such that $C = C_X \cap L$, we have $I_X = I_{X|L} = I_{X|C}$ by (1.2), which holds true by Theorem 1.2 because K_X has the projection property. Therefore, I_X is not affine on J in either case and thus I_X is strictly convex. □

References

[1] Akopyan, A. and Vysotsky, V.: Large deviations of convex hulls of planar random walks and Brownian motions. Accepted in *Ann. H. Lebesgue*, (2021). arXiv:1606.07141v4

[2] Barndorff-Nielsen, O.: Information and exponential families in statistical theory. Reprint of the 1978 original. *John Wiley & Sons, Ltd.*, Chichester, 2014. MR-3221776

[3] Bahadur, R.R. and Zabell, S.L.: Large deviations of the sample mean in general vector spaces. *Ann. Probab.* **7**, (1979), 587–621. MR-0537209

[4] Dembo, A. and Zeitouni, O.: Large Deviations Techniques and Applications. Corrected reprint of the second edition. *Springer-Verlag*, Berlin, 2010. MR-2571413

[5] Lifshits, M.A.: Large deviation principle: processes and empirical distributions (in Russian). *St. Petersburg State University*, St. Petersburg, 2002. https://sites.google.com/site/mlprobability/home10/ml09/Mall4.pdf

[6] Mogulskii, A.A.: Large deviations for processes with independent increments. *Ann. Probab.* **21**, (1993), 202–215. MR-1207223

[7] Rassoul-Agha, F. and Seppäläinen, T.: A Course on Large Deviations with an Introduction to Gibbs Measures. *American Mathematical Society*, Providence, RI, 2015. MR-3309619

[8] Rockafellar, R.T.: Convex Analysis. *Princeton University Press*, Princeton, NJ, 1970. MR-0274683

[9] Widder, D.V.: The Laplace Transform. *Princeton University Press*, Princeton, NJ, 1941. MR-0005923

Acknowledgments. I thank the anonymous referee for comments and suggestions.